

The Behrens-Fisher Problem: An Empirical Likelihood Approach

Lauren Bin Dong

*Department of Economics, University of Victoria
Victoria, B.C., Canada V8W 2Y2*

Revised, November, 2004

Abstract

A new theoretical solution to the Behrens-Fisher (BF) problem is developed using the maximum empirical likelihood method. The sampling properties of the empirical likelihood ratio (ELR) test for the BF problem are derived using Monte Carlo simulation for a wide range of situations. A comparison of the sizes and powers of the ELR test and the Welch-Aspin test is conducted for a special case of small sample sizes. The empirical results indicate that the ELR test for the BF problem has good power properties.

Keywords: Behrens-Fisher problem, empirical likelihood, Monte Carlo simulation, testing for normality, size and power

JEL Classifications: C12, C15

Author Contact:

Lauren Dong, Statistics Canada; e-mail: Lauren.Dong@statcan.ca; FAX: (613) 951-3292

1 INTRODUCTION

Testing the equality of the means of two normal populations when the variances are both unknown, and not known to be equal, is called the Behrens-Fisher (BF) problem. The Behrens-Fisher problem has been well known since the early 1930's. One reason for its fame is that it can be proven that there is no exact solution satisfying the classical criteria for good tests. That is, every invariant rejection region of fixed size for the problem must have some unpleasant properties (Zaman, 1996, p. 246). First-best solutions that are uniformly most powerful and invariant either do not exist or have strange properties. We need to look for second-best solutions.

In the literature associated with the Behrens-Fisher problem, there have been quite a few "solutions" proposed since the 1930's. For example, Fisher (1935 and 1941), Welch (1947), Aspin (1948), Cochran and Cox (1950), Qin (1991) and Jing (1995) have all suggested different solutions. Lee and Gurland (1975) presented a detailed comparison of several selected tests and proposed a refined solution to the BF problem. These various solutions can be classified into two categories: approximate tests and asymptotic tests.

The purpose of this paper is as follows. First, we derive a new theoretical solution to the Behrens-Fisher problem. The approach that we take is based on the relatively new nonparametric technique, the maximum empirical likelihood (EL) method (Owen, 2001). The way that we exploit the data and information in this paper is distinct from any previous research in both the areas of EL and the BF problem. Our EL approach makes most efficient use of the information available among the studies. Our EL approach ties together nicely the estimation and testing procedures for the equality of two means and/or the distribution of the underlying population. Second, we provide sampling properties for the ELR test using Monte Carlo simulations. The size and the power of the ELR test in finite samples are provided for a range of situations. The results indicate that the EL approach to the BF problem is both efficient and easily applicable. Third, we conduct a power comparison of the empirical likelihood ratio (ELR) test and the Welch-Aspin (WA) (1947) test for the BF problem. The results are interesting, as the ELR test is an asymptotic test, while the WA test is an approximate test in small samples.

The EL method makes use of likelihood function and the moment conditions of the data without assuming a specific parametric form for the underlying distribution. It has the flexibility to incorporate various information about the data into the approach. This leads to the efficiency of the method. Applying the EL method to the Behrens-Fisher problem provides us with a new view of the EL method and it demonstrates that the EL method is a useful tool for solving various statistical problems.

There have been two empirical likelihood type (EL-type) approaches to the BF problem in the literature; that of Qin (1991) and that of Jing (1995). Details of these two approaches are provided in Section 2.2. The EL approach of this paper is quite distinct from those of Qin and Jing in two respects. First, our way of using the data is different. Based on the knowledge that two samples are independently drawn from two different distributions that are in the same family, normal for the BF problem, we transform the first data set S_1 into a data set S_a that has the same theoretical distribution as the second data set S_2 . Then we combine the transformed data set S_a and the second data set S_2 into a full data set S such that the data have a unique distribution. Second, we exploit the data information in an more efficient way. Based on the knowledge that the first five moments of the data exist, we use these five moment equations as constraints to set up the EL method. Then, we apply the usual EL approach to the full data set S . The ELR test for the BF problem is then constructed. Details of this are provided in Section 2.3.

The outline of this paper is as follows. Section 2 provides a brief review of the conventional solutions to the BF problem in the literature. The Welch-Aspin (WA) test and the approaches of Qin (1991) and Jing (1995) are discussed in this section. Section 3 discusses the design of the new EL approach to the BF problem. It provides the ELR test and the EL-type Wald test for the BF problem. Section 4 presents some Monte Carlo experiments and the associated results for the new EL approach. The sizes and the power of the ELR test in finite samples are analyzed in detail across a broad range of situations. The comparison of the ELR test and the WA test is given in this section. Section 5 provides a summary and some conclusions.

2 SOLUTIONS TO THE BF PROBLEM

Suppose $S_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}, i = 1, 2$ are two sets of the observed values of two random samples independently drawn from the normal populations $N(\mu_i, \sigma_i^2), i = 1, 2$, respectively. Let

$$\rho^2 = \frac{\sigma_2^2}{\sigma_1^2}, \quad (1)$$

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad (2)$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \quad (3)$$

$$C_0 = \frac{\frac{\sigma_1^2}{n_1}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \frac{1}{1 + \rho^2 \frac{n_1}{n_2}} \quad (4)$$

denote the variance-ratio, sample means, sample variances, and the conventional parameter used in the literature. The problem in hand is to construct a test of the null hypothesis, $H_0 : \mu_1 = \mu_2$, against the alternative hypothesis, $H_a : \mu_1 \neq \mu_2$. Various tests have been developed by Fisher (1935), Welch and Aspin (1947), Cochran and Cox (1950) and others. The critical regions corresponding to each of these conventional tests have the general form of:

$$|v| > V(\hat{C}), \quad (5)$$

where $v = (\bar{x}_1 - \bar{x}_2) \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^{-1}$, $\hat{C} = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^{-1}$. $V(\hat{C})$ is a function of \hat{C} and the preassigned nominal significance level α . The distribution of v is no longer Student- t when the two variances are not known to be equal. The critical regions depend on the variance parameters $\sigma_i^2, i = 1, 2$ and sample sizes (n_1, n_2) . These conventional solutions to the BF problem involve utilizing various means to approximate the distribution of the variable v and to control for the size distortion of the tests. The formulars and methods of approximation are too complex to repeat here. Further details and discussion can be found in Lee and Gurland (1975). We have chosen Welch-Aspin test as an example for a brief review.

2.1 Welch-Aspin Test

Welch (1947) and Aspin (1948) independently developed a higher order approximation, $V_{wa}(\hat{C})$, to the distribution of $V(\hat{C})$ in the terms of up to f_i^{-2} and f_i^{-4} , where $f_i = n_i - 1$

and $i = 1, 2$. Their approach is referred to as the Welch-Aspin (WA) test. The WA test is a highly efficient solution to the BF problem (Weerahandi, 1987). The actual sizes of the Welch-Aspin test in small samples are very close to the nominal significance levels. For example, at a nominal level of 5%, the size of the WA test lies between 4.98% and 5.02% for $(n_1, n_2) = (7, 7)$. However, the functional form of $V_{wa}(\hat{C})$ is lengthy, involving infinite series, and it is difficult to work with (Lee and Gurland, 1975).

Lee and Gurland (1975) provided detailed comparisons of various tests that were proposed by Fisher (1935), Cochran and Cox (1950), Welch (1937), Welch-Aspin (1947) and others. At the last stage of their comparison, all of the tests were eliminated from their tables, except the Welch-Aspin test because the WA test has very accurate size and good power properties. In addition to this, Lee and Gurland proposed a refined test, which we call the LG test, using two techniques: (i) a simple functional form to approximate the $V(C)$ function; and (ii) a T-transform to accelerate the convergence of the size and the power functions. The method in their research involves solving a minimization problem of the squared difference between the size function of the test and the preassigned nominal level, and the T-transform is not trivial. Lee and Gurland provided refined results that were very close to the results of the WA test. For the reasons mentioned above, we have chosen the WA test and to cite the results of the WA test from Lee and Gurland for our comparison with the ELR test for the case of $(n_1, n_2) = (7, 7)$ at $\alpha = 5\%$. Of course, we also consider other sample sizes and nominal significance levels when considering the ELR test. A detailed discussion is given in Section 3.

2.2 Approaches of Qin and Jing

Jing (1995) proposed a nonparametric version of the EL approach to the two-sample problem. In his paper, he showed that the nonparametric version of Wilks' theorem holds in the two-sample problem and that the solution is Bartlett correctable. The empirical likelihood ratio statistic has a limiting chi-squared distribution with an error of order $O(n^{-1})$ and with the Bartlett correction, the error is reduced to the order of $O(n^{-2})$, where n is the smaller one of the two sample sizes. A special case of his approach is a particular solution to the Behrens-Fisher problem.

The focus of Jing (1995) was primarily on the coverage accuracy and the Bartlett cor-

rection of the test. The approach he used works only when the null hypothesis is true. This means that the restriction that two means are equal must be binding, and the Lagrangian multiplier for the constraint must not equal zero. If the restriction is not binding, the solution of his approach reduces to a situation where the probability parameter $p_i = 1/n$, and the estimated empirical mean of the data is just the sample mean. His approach does not appear to offer the potential for dealing with issues relating to power in the context of an ELR test.

Qin (1991) generalized Owen's empirical likelihood to a two-sample problem in which one sample is assumed to come from a distribution that is unknown and the other sample is assumed to come from a known distribution specified upto a parameter, $g_\theta(x_2)$. Qin's approach is a semi-parametric one by combining the EL method and the parametric likelihood method to the two-sample problem. Consider the following assumptions:

- $\mu_2 = \mu(\theta)$ is differentiable and $\mu'(\theta) \neq 0$
- the known density function $g_\theta(x_2)$ is differentiable three times with respect to θ
- $\int g_\theta(x_2)dx_2$ is twice differentiable
- The Fisher information matrix, $I(\theta) = E[\frac{\partial \log g_\theta}{\partial \theta}]^2$, satisfies $0 < I(\theta) < \infty$
- $|\frac{\partial^3}{\partial \theta^3} \log g_\theta(x_2)|$ is bounded.

Under these assumptions, there exists an EL estimator $\hat{\theta}$ of the mean μ_1 that is more efficient than the sample mean \bar{x}_1 . Qin proved that the empirical likelihood ratio has a limiting distribution of $\chi_{(1)}^2$ under the null hypothesis $\mu_1 = \mu_2 = \mu(\theta)$. The coverage accuracy is of the order $O_p(n^{-\frac{1}{2}})$. The connection between the two data sets is brought in only through the single restriction that the two means are equal.

The EL approach developed in this paper provides a new theoretical solution to the BF problem and hopefully it can overcome some of the shortcomings mentioned above. The focus of the following sections is on deriving the ELR test for the BF problem and simulation of the sampling properties for the ELR test in the context of solving the Behrens-Fisher problem.

3 THE EL APPROACH

3.1 ELR Test

Suppose $S_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$, $i = 1, 2$, are the two data sets we have. They are independently drawn from two normal populations $N(\mu_i, \sigma_i^2)$, $i = 1, 2$. The parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ are unknown. n_i 's are the sample sizes of the data sets. Without losing generality, we assume that $n_1 \geq n_2$. Solving the BF problem involves constructing a test for the hypothesis that the two population means equal when the two variances are not known to be equal. The steps of our approach are as follows.

First, we transform the data set S_1 , into a data set that has the same theoretical distribution as the data set S_2 . The transformation has the following form:

$$t_j = (x_{1j} - \mu_1)(\rho^2)^{\frac{1}{2}} + \mu_2, \quad (6)$$

where $x_{1j} \in S_1$ and $\rho^2 = \sigma_2^2/\sigma_1^2$. Thus, $t_j \sim N(\mu_2, \sigma_2^2)$, $j = 1, 2, \dots, n_1$.

The second data set S_2 remains unchanged. We denote: $t_{n_1+j} = x_{2j}$, where $x_{2j} \in S_2$, for $j = 1, 2, \dots, n_2$. Let $n = n_1 + n_2$. Then, the full data set $S = \{t_1, t_2, \dots, t_n\}$ is of size n and has a distribution of $N(\mu_2, \sigma_2^2)$.

The parameter $C_0 = 1/(1 + \frac{\sigma_2^2 n_1}{\sigma_1^2 n_2})$ is the one that is used in the literature of the BF problem, (*e.g.*, Welch (1947), Lee and Gurland (1975)), where the σ_i^2 's are the population variance parameters, $i = 1, 2$. Obviously, $C_0 \in (0, 1)$, and it is a smooth function of the variance ratio parameter, ρ^2 , and the ratio of the sample sizes, n_1/n_2 . In our study, we keep the ratio of the sample sizes constant and allow the variance ratio to vary such that the corresponding values of C_0 cover well the range of $(0, 1)$.

The next step is to apply the EL approach to the full data set S . A probability parameter p_j is assigned to each data point t_j . The empirical likelihood function for the full data set is formed as $\prod_{j=1}^n p_j$. The approach of the empirical maximum likelihood method is to maximize the empirical likelihood function subject to the probability constraints, $0 < p_j < 1$ and $\sum_{j=1}^n p_j = 1$, and some information constraints.

The information that we have is that the data are independent and are normally dis-

tributed with a mean μ_2 and a variance σ_2^2 . We choose to use the ratio of the variances as one of the parameters, so the parameter vector becomes $\theta = (\mu_1, \mu_2, \rho^2, \sigma_2^2)'$. In addition, we choose to use the first five unbiased raw moment equations as the information constraints, *i.e.* $m = 5$, so that the number of the moment equations, m , is greater than the number of the parameters, p . As we know, if $m < p$, the system is under-identified; and if $m = p$, the EL solution becomes exactly the solution from the method of moments. Therefore, five moment equations are necessary and sufficient for the EL method to be effective. We denote the information constraints as $E_p h(t_j, \theta) = 0$, which have the following form:

$$\sum_{j=1}^n p_j t_j - \mu_2 = 0 \quad (7)$$

$$\sum_{j=1}^n p_j t_j^2 - (\mu_2^2 + \sigma_2^2) = 0 \quad (8)$$

$$\sum_{j=1}^n p_j t_j^3 - (\mu_2^3 + 3\sigma_2^2 \mu_2) = 0 \quad (9)$$

$$\sum_{j=1}^n p_j t_j^4 - (\mu_2^4 + 6\sigma_2^2 \mu_2^2 + 3\sigma_2^4) = 0 \quad (10)$$

$$\sum_{j=1}^n p_j t_j^5 - (\mu_2^5 + 10\sigma_2^2 \mu_2^3 + 15\sigma_2^4 \mu_2) = 0. \quad (11)$$

The corresponding Lagrangian function is:

$$G = n^{-1} \sum_{j=1}^n \log p_j - \eta \left(\sum_{j=1}^n p_j - 1 \right) - \lambda' \sum_{j=1}^n p_j h(t_j, \theta), \quad (12)$$

where λ is a $m \times 1$ vector, which together with the scalar η are the Lagrangian multipliers.

The optimization problem is to maximize the Lagrangian function with respect to p_j 's, λ , and θ . Keeping λ and θ fixed, applying the first order condition with respect to and the probability parameter constraints on p_j 's, we find that η takes the value of unity and the p_j 's can be expressed as functions of λ and θ :

$$p_j = n^{-1} (1 + \lambda' h(t_j, \theta))^{-1}, \quad j = 1, 2, \dots, n \quad (13)$$

Substituting this information back into the Lagrangian function, we get a optimization problem over a reduced number of unknowns, λ and θ .

Deriving the first order conditions of the Lagrangian function with respect to the parameter vector θ requires some special attention. We note that the values in the first portion of the data are functions of the unknown parameters. The first derivative of the data with respect to the parameters involves the following terms:

$$\frac{\partial t_j}{\partial \theta} = (-\rho, 1, (x_{1j} - \mu_1) \frac{1}{2} (\rho^2)^{-1/2}, 0)', \quad j = 1, 2, \dots, n_1, \quad (14)$$

where $x_{1j} \in S_1$. Taking into account this information, the actual first order conditions for the Lagrangian function with respect to the four unknown parameters $(\mu_1, \mu_2, \rho^2, \sigma_2^2)'$ can be represented as follows:

$$\begin{aligned} & \sum_{j=1}^{n_1} p_j (\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4) (\rho^2)^{1/2} = 0 \\ & \sum_{j=1}^{n_1} p_j (\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4) \\ & \quad - \sum_{j=1}^n p_j (\lambda_1 + 2\lambda_2 \mu_2 + 3\lambda_3 (\sigma_2^2 + \mu_2^2) + 4\lambda_4 (3\sigma_2^2 \mu_2 + \mu_2^3) + 5\lambda_5 (3\sigma_2^4 + 6\sigma_2^2 \mu_2^2 + \mu_2^4)) = 0 \\ & \sum_{j=1}^{n_1} p_j (\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4) (t_j - \mu_1) \frac{1}{2} (\rho^2)^{-1/2} = 0 \\ & \sum_{j=1}^n p_j (\lambda_2 + 3\lambda_3 \mu_2 + 6\lambda_4 (\sigma_2^2 + \mu_2^2) + 10\lambda_5 (3\sigma_2^2 \mu_2 + \mu_2^3)) = 0. \end{aligned}$$

Putting these four first order conditions and the five moment equations together, we get a system of nine nonlinear equations. We solve this system using the nonlinear equation solver ‘‘Eqsolve’’ in the Gauss package (Aptech Systems, 2002).

We denote $\hat{\lambda}$ and $\hat{\theta}$ as the EL estimators, the solution from the system, for the Lagrangian multiplier vector and parameter vector. Then, we obtain the EL estimators for the probability parameters \hat{p}_j 's using the formula (13). The estimated maximum value of the likelihood function is obtained as $L(\hat{F}^u) = \prod_{j=1}^n \hat{p}_j^u$, where u stands for the unconstrained model.

In order to construct an ELR test for the BF problem, we also need to have the solutions to the constrained case. With the null hypothesis: $H_0 : \mu_1 = \mu_2$, we simply substitute the restriction $\mu_1 = \mu_2$ into the unconstrained case to get the constrained case, where the parameter vector becomes $\theta = (\mu, \rho^2, \sigma_2^2)'$ which is of dimension three, with $\mu = \mu_1 = \mu_2$.

The five moment equations, $E_p h^c(t_j, \theta)$, are:

$$\sum_{j=1}^n p_j t_j - \mu = 0 \quad (15)$$

$$\sum_{j=1}^n p_j t_j^2 - (\mu^2 + \sigma_2^2) = 0 \quad (16)$$

$$\sum_{j=1}^n p_j t_j^3 - (\mu^3 + 3\sigma_2^2 \mu) = 0 \quad (17)$$

$$\sum_{j=1}^n p_j t_j^4 - (\mu^4 + 6\sigma_2^2 \mu^2 + 3\sigma_2^4) = 0 \quad (18)$$

$$\sum_{j=1}^n p_j t_j^5 - (\mu^5 + 10\sigma_2^2 \mu^3 + 15\sigma_2^4 \mu) = 0. \quad (19)$$

where c stands for constrained. Similarly, we have:

$$p_j^c = n^{-1}(1 + \lambda' h^c(t_j, \theta))^{-1}, \quad j = 1, 2, \dots, n \quad (20)$$

The first derivative of the first portion of the data with respect to the parameters has the following form:

$$\frac{\partial t_j}{\partial \theta} = (-(\rho^2)^{1/2} + 1, (x_{1j} - \mu) \frac{1}{2} (\rho^2)^{-1/2}, 0)', \quad j = 1, 2, \dots, n_1, \quad (21)$$

where $x_{1j} \in S_1$. The first order conditions of the Lagrangian function with respect to the three parameters are:

$$\begin{aligned} & \sum_{j=1}^{n_1} p_j (\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4) ((\rho^2)^{1/2} - 1) \\ & - \sum_{j=1}^n p_j (\lambda_1 + 2\lambda_2 \mu + 3\lambda_3 (\sigma_2^2 + \mu^2) + 4\lambda_4 (3\sigma_2^2 \mu + \mu^3) + 5\lambda_5 (3\sigma_2^4 + 6\sigma_2^2 \mu^2 + \mu^4)) = 0 \\ & \sum_{j=1}^{n_1} p_j (\lambda_1 + 2\lambda_2 t_j + 3\lambda_3 t_j^2 + 4\lambda_4 t_j^3 + 5\lambda_5 t_j^4) (t_j - \mu) \frac{1}{2} (\rho^2)^{-1/2} = 0 \\ & \sum_{j=1}^n p_j (\lambda_2 + 3\lambda_3 \mu + 6\lambda_4 (\sigma_2^2 + \mu^2) + 10\lambda_5 (3\sigma_2^2 \mu + \mu^3)) = 0. \end{aligned}$$

Solving the system of the nonlinear equations of the five moment equations and the three first order conditions, we get $\hat{\lambda}^c$ and $\hat{\theta}^c$ as the EL estimates, and then, the \hat{p}_j^c 's using the

formula (20). The estimated maximum value of the empirical likelihood function under the constraint is formed as $L(\hat{F}^c) = \prod_{j=1}^n \hat{p}_j^c$.

The log likelihood ratio statistic has the form:

$$\begin{aligned} -2 \log R(F) &= -2 \log \frac{L(\hat{F}^c)}{L(\hat{F}^u)} \\ &= 2 \sum_{j=1}^n (\log(1 + \hat{\lambda}^c h^c(t_j, \hat{\theta}^c)) - \log(1 + \hat{\lambda}^u h(t_j, \hat{\theta}^u))) \end{aligned}$$

and it has a limiting distribution of $\chi_{(1)}^2$ under H_0 .

3.2 EL-type Wald Test

The EL estimator $\hat{\theta}_{EL}$ of the parameter vector θ is asymptotically efficient, and it has a limiting distribution of the following form:

$$\sqrt{n}(\hat{\theta}_{EL} - \theta_0) \xrightarrow{d} N(0, \Sigma),$$

where

$$\Sigma = [E[\frac{\partial h(y, \theta)}{\partial \theta} |_{\theta_0}] E[h(y, \theta)h(y, \theta)' |_{\theta_0}]^{-1} E[\frac{\partial h(y, \theta)}{\partial \theta'} |_{\theta_0}]]^{-1}, \quad (22)$$

and θ_0 is the true value of θ . A consistent estimator of the asymptotic covariance matrix Σ can be obtained using the EL estimator $\hat{\theta}_{EL}$:

$$\hat{\Sigma} = [E[\frac{\partial h(y, \theta)}{\partial \theta} |_{\hat{\theta}_{EL}}] E[h(y, \theta)h(y, \theta)' |_{\hat{\theta}_{EL}}]^{-1} E[\frac{\partial h(y, \theta)}{\partial \theta'} |_{\hat{\theta}_{EL}}]]^{-1}. \quad (23)$$

The estimated covariance matrix $\hat{\Sigma}$ is a 4×4 symmetric matrix. With this, we can easily obtain an EL-type Wald test for any linear restrictions of the parameters. Suppose $c\hat{\theta} - r = 0$ is a set of j linear restrictions. The EL-type Wald test has the form:

$$W = (c\hat{\theta} - r)' [c\hat{\Sigma}c']^{-1} (c\hat{\theta} - r), \quad (24)$$

and it has an asymptotic distribution of $\chi_{(j)}^2$, if the restrictions are valid.

For the BF problem, the restriction is simply $\mu_1 = \mu_2$, *i.e.*, $c = \{1, -1, 0, 0\}$, then, the

EL-type Wald test has the form of:

$$w = (\mu_1 - \mu_2)^2(a_{11} + a_{22} - 2a_{12})^{-1}, \quad (25)$$

where a_{ij} , $i, j = 1, 2$ are the elements of the $\hat{\Sigma}$ matrix. The EL-type Wald test statistic w has an asymptotic distribution of $\chi^2_{(1)}$ under the null hypothesis.

We have explored the EL-type Wald test for the BF problem using Monte Carlo simulations. Intuitively, the Wald test should be computationally easier than the ELR test since the Wald test involves solving the unconstrained case only. However, our study finds that both the computing time and computational difficulty associated with the Wald test are greater than that associated with the ELR test. One possible reason for this is that the Wald test involves computing the consistently estimated covariance matrix for the parameter estimators, and this involves three matrix inversions. For an ill behaved problem, such as the Behrens-Fisher problem, these matrix inversions may cause some difficulties. The results for the Wald test are not as good as those for the ELR test in the BF context. Therefore, we choose to present only the results of the ELR test in this study (The details of the Wald test results are available on request.)

3.3 Testing For Normality

Solving the Behrens-Fisher problem and testing for normality have close connections when using the EL method as described in the previous sections. By definition, the BF problem is the case of testing the equality of two means when two data sets are independently drawn from two *normal* distributions with general and unknown variances. It was categorized by Tukey (1954) as the fourth level problem of *normal sequences of growth in consideration in the area of comparing the typical values of two populations with the aid of a sample drawn from each* (Tukey, 1954, p. 713).

The EL approach that we described in Section 3.1 provided a new solution to the Behrens-Fisher problem. Further more, suppose these two samples were drawn from two different distributions that are from the same family, and we have interest in making sure that this distribution family is normal. Then, our approach is perfectly suitable for us to conduct an ELR test for *normality* of the underlying distributions of the two data sets.

As described in Dong and Giles (2004), testing for the validity of the moment conditions provides a way of testing for the underlying distribution, normality in this case.

Consider two data sets $S_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$, $i = 1, 2$. Suppose they are drawn independently from two populations $F(\mu_i, \sigma_i^2)$ of the same family, where F is not known and is not necessarily normal. As described in section 3.1, we apply the EL approach to the full data set S . If the underlying distribution is normal, then the five moment equations constraints hold true, and the maximized likelihood function is $L(\hat{F}^c) = \prod_{j=1}^n \hat{p}_j$. If the hypothesis is not true, then the maximum value of the likelihood function is $L(F^u) = n^{-n}$. The log empirical likelihood ratio statistic for testing for normality is of the form:

$$-2 \sum \log n\hat{p}_j = 2 \log(1 + \hat{\lambda}'h(t_j, \hat{\theta})). \quad (26)$$

It has a limiting null distribution of $\chi_{(1)}^2$, where the number of degrees of freedom equals the number of moment equation, five, less the number of parameters, four.

Usually, we may be interested in using this technique to test for normality of the two underlying populations of the same distribution family. If we are satisfied with the results and accept the null hypothesis that the underlying populations are normal, then, we can continue the process to solve the Behrens-Fisher problem. However, the first step of testing for normality introduces a pre-testing situation that the size and power of the subsequent test in the procedure may be altered, as described by Giles and Giles (1993). It is well known that sequential testing strategies can result in size and power distortions if the tests in question are not independent of each other. This issue is not explored further in this study.

Alternatively, we can avoid the pre-testing issue by conducting an ELR test for normality and an EL-type Wald test for the Behrens-Fisher problem simultaneously. Using the unrestricted model and applying the EL approach described in Section 3.1, we obtain the EL estimate of the parameter vector, $\hat{\theta}$, and the estimates of the probability parameter, \hat{p}_i 's. With the \hat{p}_i 's, we can implement the ELR test for normality for the underlying distribution as we described in Dong and Giles (2004). In the meantime, without being influenced by the testing for normality, we can compute the consistent estimate of the covariance matrix of $\hat{\theta}$ to perform the EL-type Wald test for the Behrens-Fisher problem as we described at the beginning of this section. If we accept the null hypothesis of the first test that the underlying distribution of the two populations is normal, then the Wald test is well-founded.

This alternative approach seems to be able to effectively avoid the pre-testing issue, and it may provide better results. We will explore this alternative in future research.

3.4 Advantages of the EL Approach

The empirical likelihood approach to the BF problem described in this paper utilizes more information from the data sets than do the approaches suggested by other authors. In particular, we exploit the obvious relationship between the two data sets that they both are drawn from normal distributions. We make use of the first five unbiased moment equations of the data sets. In addition, we utilize the empirical likelihood functions. These techniques allow the EL approach of our design to reach a higher asymptotic efficiency for estimation and higher testing power for the new solution to the BF problem. If the ELR test had a well controlled size and good power, then we would say that the ELR test is a good solution to the BF problem and the EL approach is valuable. Monte Carlo simulations that are described in the following section provide an extensive analysis of the sampling properties of the new ELR test.

4 MONTE CARLO EXPERIMENTS

For the Behrens-Fisher problem, the Monte Carlo experimental design is to apply the empirical likelihood method to test for the equality of two means without the knowledge of the variances of the underlying populations.

Two random samples are generated independently from two normal distributions: $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with the true values of the parameters $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)' = (1, 1, 1, \sigma_2^2)'$, where σ_2^2 varies with the variance ratio parameter $\rho^2 = \sigma_2^2/\sigma_1^2$. The true value of the parameter ρ^2 changes according to the values of $\{0.1, 0.5, 1, 2, 10\}$. When $\rho^2 = 1$, the two variances equal; when ρ^2 is farther apart from one, the two variances are further different from each other where the BF problem becomes severe. The sample size pair ranges from (20, 10) to (250, 125). The ratio of the sample sizes of the two data sets is kept constant, $n_1/n_2 = 2$, for simplicity. These settings lead to the corresponding values of the parameter used in the

literature $C_0 = 1/(1 + \frac{\sigma_2^2}{\sigma_1^2} \frac{n_1}{n_2})$, to vary in accordance of $\{0.83, 0.5, .33, 0.2, 0.048\}$ which has a good coverage of the region $(0, 1)$. Therefore, our experiments have a good coverage of the parameter space. The number of replications is set to be 5,000.

In conducting the power comparisons, the non-centrality parameter is defined as:

$$\delta = \frac{\mu_1 - \mu_2}{(\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}}. \quad (27)$$

The null hypothesis is true when $\delta = 0$. The alternative hypothesis is when $\delta \neq 0$. We have considered are the cases of δ varies according to the values of $\{1, 1.5, 2, 2.5, 3, 4\}$.

One note we should bear in mind is that, in contrast to the non-centrality parameter used in Lee and Gurland (1975), the parameter δ in our design does not depend on the sample size pair (n_1, n_2) . Therefore, we are able to disentangle the effect upon the power of the ELR test that is arising from the non-centrality parameter δ , or that from sample size pair changes. The log empirical likelihood ratio statistic depends on the parameter ρ^2 and the sample size ratio n_1/n_2 . Thus, the size and the power of the ELR test are also functions of these parameters. However, the asymptotic distribution of the ELR test statistic does not depend on these nuisance parameters. The ELR test statistic has an asymptotic distribution of $\chi_{(1)}^2$ under the null hypothesis.

5 EMPIRICAL RESULTS

5.1 Sampling properties

Table 1 in the Appendix presents the computed sizes and the size-adjusted critical values for the ELR test in solving the Behrens-Fisher problem at four nominal significance levels: $\alpha = \{10\%, 5\%, 2\%, 1\%\}$. These four levels are used so as to have a close look at the behavior of the ELR test at the tail.

- The actual size of the ELR test is relatively large comparing to the nominal significance level. The size has an appropriate trend to converge to the correct nominal significance

level when the sample sizes increase, given a value of ρ^2 . For example, the size changes from 12.14% to 10.66% when the sample pair varies from (20, 10) to (250, 125) at the nominal significance level of 5% when $\rho^2 = 0.1$. We stopped at $n = (250, 125)$ due to computing difficulty. It would be worthwhile to provide further empirical evidence to illustrate the size convergence as the sample size increases in future research.

- The size of the ELR test is not very sensitive to the changes in the sample size pair at the nominal significance level of 10%.
- The size distortion of the test declines as the value of the parameter ρ^2 decreases, holding other factors fixed. For instance, the size changes from 27.36% to 12.14% when ρ^2 moves from 10 to 0.1 for small sample sizes $(n_1, n_2) = (20, 10)$ at $\alpha = 5\%$. The size distortion is the worst when $\rho^2 = 10$ however it is still within the range of our expectations for the ELR test, based on our computational experience.

Tables 2 to 4 provide the full analysis of the power of the ELR test across a range of values for the non-centrality parameter δ , the parameter ρ^2 , and the sample size pair (n_1, n_2) . The size-adjusted critical values are used to compute the powers of the test in every cases. These values allow us to evaluate the power of the ELR test at the actual significance levels. The findings are as follows.

- First, the power increases as the parameter δ increases, given the values of ρ^2 , α , and (n_1, n_2) . That is the power of the test is higher when the hypothesis is further away from the truth. For example, the power increases from 25.20% to 95.46% when the parameter δ varies from 1 to 4 at $\rho^2 = 0.1$ and $\alpha = 5\%$ for the sample size pair as small as $(n_1, n_2) = (20, 10)$. These are very encouraging results. They show that the ELR test has good power properties for quite small samples.
- Second, the power of the test increases with the sample sizes. That is, the test is more powerful when we have more information in hand. For instance, the power increases from 49.26% to 99.94% when the sample size pair increases from (20, 10) to (250, 125) at $\rho^2 = 0.1$, $\alpha = 5\%$, and $\delta = 1.5$.
- Third, the power is higher when the value of ρ^2 is farther away from unity, given the values of the parameters of (n_1, n_2) , α , and δ . For example, the power changes from 30.16% to 90.28% when ρ^2 changes from 1 to 0.1; the power increases to 66.30% when

the parameter ρ^2 moves from 1 to 10, holding $\alpha = 5\%$, $\delta = 2.5$, and $(n_1, n_2) = (20, 10)$ fixed. The rationale behind this result is as follows. (i) When $\rho^2 = 1$, the variances of the two populations are unknown but equal, where the BF problem vanishes. The test for $H_0 : \mu_1 = \mu_2$ reduces to the usual t test. (ii) When ρ^2 deviates from unity, the sign of the Behrens-Fisher problem shows up, the farther away ρ^2 is from unity, the severe the BF problem becomes; in this case, the ELR test becomes more powerful. In another words, under the alternative hypotheses, when the variances of the two populations are unknown and unequal, the ELR test has high power. The result tells us that the ELR test is able to capture the specific information that the samples were drawn from two populations with different variances that are unknown in the BF problem.

Overall, the power results are quite acceptable and the ELR test is recommended to be applied to the BF problem.

5.2 Comparison of the WA and the EL Tests

The WA test is designed for small samples. As discussed in Section 2.1, the solution provided by the Welch-Aspin test results from directly solving the size function equation

$$P(|v| > V(\hat{C})) = \alpha \quad (28)$$

and approximating $V(\hat{C})$ function in the terms of, and up to, f_i^{-4} , where $f_i = n_i - 1$ is the number of degrees of freedom left for each data set, and $i = 1, 2$. We cite the experimental results of the WA test from Lee and Gurland (1975). The results were specifically for the case of $(n_1, n_2) = (7, 7)$, $\alpha = 5\%$, and $C_0 = \{0.1, 0.2, 0.3, 0.4, 0.5\}$. The sample size is very small $(n_1, n_2) = (7, 7)$, but the test has well controlled size and good power.

In this section, we conduct a comparison of the sampling properties between the ELR test and the WA test for this small sample size case and results are provided in Table 5 in the appendix.

To make it comparable, the parameter ρ^2 in the ELR test takes the values of $\{9, 4, 2.33, 1.5, 1\}$ in corresponding to the values of the parameter C_0 used by Lee and Gurland (1975). The non-centrality parameter δ takes the values of $\{0.378, 0.7559, 1.1339, 1.5119\}$ in corresponding to

the non-centrality parameters of the WA test which changes values according to $\{1, 2, 3, 4\}$. The two non-centrality parameters have the following relationship: $\delta = \delta_0/\sqrt{7}$.

From the table, we see that the ELR test, unfortunately, performs poorly for this very small-sample situation. We wish to make more comparisons of these two tests for different samples. However, to our knowledge, there is no other published information available to allow a comparison between the ELR and the WA test for sample size that is greater than seven.

For the sample size pair as small as $(n_1, n_2) = (7, 7)$, the actual size of the ELR test obviously exceeds the nominal significance level but it is still within the usual size range of the ELR test. The size distortion of a test is the difference between the actual size of the test and the nominal significance level. Therefore, the size distortion of the ELR test is large.

We compute the power of the ELR test at its actual size level of 5% by using the simulated size-adjusted critical values. The power comparison is made at the same actual size level 5%. The result shows that the power of the ELR test for such a small sample size is low and it is inferior to that of the WA test. It is unfortunate that the ELR test is not able to show its merits for this extremely small sample size pair.

The ELR test for the Behrens-Fisher problem is an asymptotic test. The power performance is acceptably good when the sample size pair is as small as $(n_1, n_2) = (20, 10)$. However, we would not expect it to perform very well when the sample size pair is extremely small, such as $(n_1, n_2) = (7, 7)$. The power results from Monte Carlo experiment for this specific case (but with various significance levels) are presented in Table 6.

Although this particular comparison between the ELR test and the WA test is rather disappointing, it must be kept in perspective. First, we are comparing our asymptotic test with one which is explicitly designed for small samples. Second, this comparison involves a particularly small sample size.

The full experimental results for the ELR test solving the Behrens-Fisher problem that are presented in Tables 2 to 5 show that the EL method is able to solve the BF problem, and that the ELR test has good power properties over a wide range of realistic situations.

6 COMPUTATIONAL ISSUES

The computing work associated with our EL approach in solving the BF problem is challenging. As Owen (2001) stated, it is computationally challenging to optimize a likelihood function of either parametric or empirical type over some nuisance parameters with other parameters held fixed at test values.

The BF problem is well known to be very difficult to solve. This type of difficulty in optimizing empirical likelihood functions is especially clear. (i) The BF problem involves two nuisance parameters; (ii) Along the boundary of the parameter ρ^2 , empirical likelihood function is ill-behaved according to our experiment.

There are two possible reasons for this difficulty. One reason is historical. Any solution to the Behrens-Fisher problem must have some unpleasant properties (Zaman, 1996, p. 246). The ELR test, like other tests, must have some unfavorable features over certain areas in the parameter space. Our work proves this is true.

The second possible reason comes from the design of the empirical likelihood approach. The nature of the EL method is that, in the neighborhood of the solution, the gradient matrix associated with the moment constraints will approach an ill-behaved state of being less than full rank (Mittelhammer *et al.*, 2003). This occurs by design because the basic rationale of the EL method is to modify the sample weights such that the over-identified m empirical moment equations can be satisfied in order to solve for the unique solutions of the p unknowns, where $m > p$. This creates instability in gradient-based constrained optimization algorithms regarding the representation of the feasible spaces and feasible directions for such problems.

Mittelhammer *et al.* (2003) used a "concentrating-out" technique that utilized a non-linear system procedure (NLSYS) and the Nelder-Mead (1965) method to achieve their computational purpose. These techniques worked well for their project. We have explored the possibility of using these techniques in solving the Behrens-Fisher problem. However, the Nelder-Mead method did not work well in our situation.

The approach we used in this paper is so-called "direct solve" method. We directly solve the non-linear system of the moment equations and the first order conditions with respect to the parameters. The non-linear equation solving procedure, Eqsolve, in the Gauss package

is employed. The numerical solutions are acceptably good.

In the process of the Monte Carlo simulations, there are a few samples drawn from the underlying distributions that can not solve the nonlinear system in computing the power of the ELR test. This reflects a typical example of the potential infeasibility that happens in implementing the EL method, in practise.

When this happens, we reuse the sample that did not work for the Eqsolve procedure by altering the initial values of the parameters and then implementing the procedure again until the solution is found. In finding a new vector of initial values for the parameters, we apply the essential idea of the Differential Evolution method.

Suppose $\hat{\theta}_0$ is the vector of unsuccessfully estimated values of θ . The random search direction is formed of the difference between two random vectors $\theta_1 - \theta_2$. The new initial value for the parameter vector is then in the form: $\theta_a = \hat{\theta}_0 + s(\theta_1 - \theta_2)$, where s is a step size taking values of $\{0.4, 1, 2\}$. This new initial vector of parameter values is then feeded to the Eqsolve procedure to search for the global maximum.

These techniques worked very well in keeping the samples that we have generated. By doing so we effectively avoid throwing away data sets casually and we have practically avoided the problem of selection bias. Therefore, we can claim that we have approximated the exact distribution for the ELR test statistic in finite samples.

As a result, the empirical size distortion of the ELR test is effectively improved using this data reuse technique comparing to the ones when we did not use this data recycle technique. However, any gain comes with certain trade off. The computing time is lengthened and it is significantly longer than the EL approach in the testing for normality in Dong and Giles (2004).

To give an indication of the extent of the difficulty it involved in solving the BF problem using the ELR method, we provide some examples of the computing times that it took in the Monte Carlo experiments.

The computing time is longer when the parameter ρ^2 is away from unity. It takes about 10 hours to compute the empirical size for the ELR test when $\rho^2 = 2$, $(n_1, n_2) = (20, 10)$, and the number of replications is 5,000 using a Pentium 4, 2.4 GHZ PC. It is also very difficult in computation when the null hypothesis is not true, e.g. when the non-centrality

parameter $\delta \neq 0$. For example, in computing the power of the ELR test, it takes the same machine around 27 hours for the case of $(n_1, n_2) = (20, 10)$, $\rho^2 = 0.5$, $\delta = 1$, the number of replications is 5,000.

In conclusion, although it is difficult in computing the exact size and power of the ELR test for the BF problem, our EL approach using the Eqsolve algorithm works well for the BF problem. The empirical results are sound.

7 SUMMARY AND CONCLUSIONS

We have developed an new theoretical approach using the EL method to solve the Behrens-Fisher problem in this paper. The fact that the EL method is able to solve the BF problem is important. It shows the flexibility of the EL method in solving various problems in statistics and econometrics.

A full range of Monte Carlo experiments are conducted to provide the analysis of the sampling properties of the ELR test. The actual sizes and the size-adjusted critical values in finite samples are simulated. The size-adjusted critical values are used to conduct the analysis of the power properties of the ELR test. The empirical results provide the evidence that the ELR test has good sampling properties across different parameter dimensions: the variance ratio parameter, the sample size pair, and the non-centrality parameter.

Generally, the size-adjusted critical values that we have presented in Table 1 are ready to be used by researchers provided that the values of the parameter ρ^2 and the sample size pair are conformable with the ones in our study.

We have noted that the computing time in solving the BF problem is significantly long. The size of the ELR test statistic, in general, is still larger than the significance levels we consider. In the future study, it would be fruitful to explore some techniques that could reduce both of the computational difficulties and the size distortion and could still maintain the good power properties of the ELR test.

Acknowledgements

This paper is based on one chapter of the author's PhD dissertation, completed in the Department of Economics, University of Victoria, in December 2003. The author is very grateful to the thesis supervisor, professor David Giles, for his timely guidance and warm support. Special thanks are also extended to Don Ferguson, Ron Mittelhammer, Min Tsao, Graham Voss and Julie Zhou for their many helpful suggestions and contributions.

Appendix: Monte Carlo Results

Table 1: Size and Size-adjusted Critical Values of the ELR Test

(n_1, n_2) :	(20, 10)	(60, 30)	(100, 50)	(250, 125)	(20, 10)	(60, 30)	(100, 50)	(250, 125)
	$\rho^2 = 0.1$				$\rho^2 = 0.5$			
10%	0.1838	0.1868	0.1832	0.1878	0.2202	0.2202	0.2266	0.1964
5%	0.1214	0.1124	0.1084	0.1066	0.1482	0.1468	0.1448	0.1218
2%	0.0712	0.0636	0.0580	0.0554	0.0924	0.0852	0.0896	0.0668
1%	0.0488	0.0428	0.0404	0.0342	0.0686	0.0584	0.0626	0.0424
<i>Size-adjusted Critical Values:</i>								
10%	4.3067	4.1787	4.0088	3.9780	5.1494	4.8877	4.9112	4.3564
5%	6.5523	6.1960	5.8025	5.6186	7.7853	7.1960	7.4334	6.1392
2%	10.7818	10.1350	9.0218	8.2500	12.4479	10.4005	11.1566	9.3809
1%	14.2266	13.2669	12.8602	10.3353	16.5913	13.6268	14.1973	11.9212
	$\rho^2 = 1$				$\rho^2 = 2$			
10%	0.2424	0.2538	0.2536	0.2332	0.2926	0.3014	0.2780	0.2448
5%	0.1698	0.1788	0.1748	0.1466	0.2136	0.2110	0.1966	0.1592
2%	0.1100	0.1126	0.1068	0.0828	0.1404	0.1422	0.1298	0.0888
1%	0.0814	0.0776	0.0710	0.0530	0.1046	0.1044	0.0966	0.0570
<i>Size-adjusted Critical Values:</i>								
10%	5.7988	5.8435	5.5645	4.7971	6.8276	6.7529	6.4580	5.1142
5%	8.9674	7.9674	7.5631	6.7822	10.8456	9.3973	9.3711	7.1188
2%	14.3630	11.8099	10.5112	9.8197	17.7402	13.1147	13.3604	10.0495
1%	18.6082	14.3278	13.6295	12.0383	22.9714	15.9200	16.7853	12.1949
	$\rho^2 = 10$							
10%	0.3564	0.3306	0.2872	0.2376				
5%	0.2736	0.2574	0.2066	0.1576				
2%	0.1892	0.1826	0.1384	0.0948				
1%	0.1516	0.1428	0.1044	0.0636				
<i>Size-adjusted Critical Values:</i>								
10%	9.25930	8.4782	6.7945	5.2401				
5%	15.0960	12.3244	10.4183	7.4935				
2%	23.8243	18.3507	16.2891	11.1926				
1%	31.6749	21.8045	21.2274	14.8243				

Notes to table:

The number of replications is 5,000.

The sample sizes are the pair (n_1, n_2) . $\rho^2 = \sigma_2^2/\sigma_1^2$.

Table 2: Power of the ELR Test for the Behrens-Fisher Problem

δ :	(n_1, n_2)											
	$\rho^2 = 0.1$						$\rho^2 = 0.5$					
	1	1.5	2	2.5	3	4	1	1.5	2	2.5	3	4
	(20, 10)											
10%	0.3826	0.6786	0.9032	0.9684	0.9750	0.9768	0.1688	0.2945	0.4942	0.7038	0.8408	0.9508
5%	0.2520	0.4926	0.7764	0.9028	0.9388	0.9546	0.0994	0.1816	0.3101	0.4562	0.608	0.8166
2%	0.1396	0.2262	0.4122	0.5364	0.5910	0.6896	0.0542	0.1028	0.1674	0.2344	0.3078	0.5022
1%	0.0906	0.1226	0.2122	0.2836	0.3298	0.4316	0.0371	0.0743	0.1226	0.1562	0.2104	0.3314
	(60, 30)											
10%	0.6086	0.9314	0.9926	0.9924			0.2870	0.5748	0.9148	0.9832	0.9906	
5%	0.4800	0.8748	0.9870	0.9846			0.1800	0.4256	0.8350	0.9692	0.9848	
2%	0.2934	0.7292	0.9724	0.9764			0.1034	0.2710	0.6926	0.9376	0.9798	
1%	0.2088	0.5964	0.9530	0.9740			0.0682	0.1812	0.5422	0.8764	0.9688	
	(100, 50)											
10%	0.6976	0.9812	0.9978				0.3180	0.7362	0.9768	0.9962		
5%	0.5992	0.9608	0.9970				0.1956	0.5988	0.9494	0.9924		
2%	0.4452	0.9074	0.9930				0.0950	0.3978	0.8850	0.9840		
1%	0.2972	0.8092	0.9848				0.0584	0.2792	0.8032	0.9742		
	(250, 125)											
10%	0.8254	0.9996					0.4216	0.9450	0.9992			
5%	0.7512	0.9994					0.3130	0.9052	0.9986			
2%	0.6308	0.9972					0.1828	0.8040	0.9962			
1%	0.5474	0.9952					0.1112	0.7026	0.9928			

Notes to table:

The number of replications is 5,000.

The sample sizes are the pair (n_1, n_2) . $\rho^2 = \sigma_2^2/\sigma_1^2$.

δ is the non-centrality parameter.

Table 3: Power of the ELR Test for the Behrens-Fisher Problem

δ :	(n_1, n_2)											
	$\rho^2 = 1$						$\rho^2 = 2$					
	1	1.5	2	2.5	3	4	1	1.5	2	2.5	3	4
	(20, 10)											
10%	0.1564	0.2170	0.3456	0.5116	0.6886	0.9050	0.1593	0.1956	0.2672	0.4400	0.6366	0.8352
5%	0.0972	0.1306	0.2078	0.3016	0.4634	0.7318	0.1044	0.1303	0.1750	0.3016	0.4612	0.6564
2%	0.0606	0.0898	0.1342	0.1872	0.2806	0.4624	0.0658	0.0950	0.1378	0.2354	0.3731	0.5456
1%	0.0420	0.0700	0.1112	0.1584	0.2312	0.3878	0.0440	0.0800	0.1240	0.2188	0.3522	0.5272
	(60, 30)											
10%	0.2076	0.4050	0.7574	0.9458	0.9846	0.9941	0.1926	0.289	0.6084	0.8770	0.9772	0.9968*
5%	0.1440	0.2878	0.6310	0.8962	0.9748	0.9914	0.1342	0.1804	0.4504	0.7666	0.9462	0.9942
2%	0.0870	0.1584	0.4192	0.7648	0.9374	0.9850	0.0869	0.0995	0.2876	0.6190	0.8794	0.9871
1%	0.0664	0.1144	0.3082	0.6614	0.8930	0.9798	0.0648	0.0716	0.2170	0.5232	0.8082	0.9683
	(100, 50)											
10%	0.2364	0.5486	0.9086	0.9918			0.1988	0.4190	0.8054	0.9774	0.9970	
5%	0.1572	0.4284	0.8452	0.9828			0.1303	0.2693	0.6634	0.9472	0.9950	
2%	0.0980	0.2770	0.7322	0.9612			0.0814	0.1380	0.4756	0.8756	0.9886	
1%	0.0698	0.1766	0.5894	0.9290			0.0620	0.0808	0.3459	0.7996	0.9776	
	(250, 125)											
10%	0.3104	0.8450	0.9964				0.2744	0.7362	0.9878	0.9996		
5%	0.2082	0.7514	0.9914				0.1844	0.6116	0.9746	0.9988		
2%	0.1152	0.6086	0.9738				0.1024	0.4572	0.9432	0.9982		
1%	0.0776	0.5048	0.9572				0.0682	0.3624	0.9084	0.9976		

Notes to table:

The number of replications is 5,000.

The sample sizes are the pair (n_1, n_2) . $\rho^2 = \sigma_2^2/\sigma_1^2$. δ is the non-centrality parameter.

Results with a star (*) are the ones that the number of replications is less than 5,000 due to computing difficulties.

Table 4: Power of the ELR tTest for the Behrens-Fisher Problem

(n_1, n_2)	$\rho^2 = 10$					
δ :	1	1.5	2	2.5	3	4
<hr/>						
(20, 10)						
10%	0.2056	0.2586	0.4852	0.7378	0.8340	0.8556
5%	0.1512	0.2084	0.4182	0.6630	0.7815	0.8088
2%	0.1046	0.1782	0.3714	0.6250	0.7482	0.7864
1%	0.0764	0.1572	0.3422	0.5928	0.7240	0.7712
<hr/>						
(60, 30)						
10%	0.1900	0.2572	0.7734	0.9894	0.9970	0.9945*
5%	0.1386	0.2060	0.7144	0.9762	0.9856	0.9872
2%	0.0798	0.1594	0.6150	0.8982	0.9277	0.9287
1%	0.0600	0.1270	0.5270	0.8072	0.8543	0.8793
<hr/>						
(100, 50)						
10%	0.1522	0.3674	0.9114	0.9995*	0.9997*	1.0000*
5%	0.0812	0.2214	0.8544	0.9987	0.9995	1.0000
2%	0.0516	0.1446	0.7945	0.9960	0.9973	1.0000
1%	0.0450	0.1296	0.7734	0.9922	0.9940	0.9968
<hr/>						
(250, 125)						
10%	0.2388	0.6006	0.9977*			
5%	0.1276	0.4794	0.9931			
2%	0.0544	0.3242	0.9851			
1%	0.0300	0.2248	0.9747			

Notes to table:

The number of replications is 5,000.

The sample sizes are the pair (n_1, n_2) . $\rho^2 = \sigma_2^2/\sigma_1^2$. δ is the non-centrality parameter.

Results with a star (*) are the ones that the number of replications is less than 5,000 due to computing difficulties.

Table 5: Size and Power Comparison of the Welch-Aspin and the ELR Tests for $(n_1, n_2) = (7, 7)$, $\alpha = .05$

δ_0	0		1		2		3		4	
C_0	V_{wa}	EL	V_{wa}	EL	V_{wa}	EL	V_{wa}	EL	V_{wa}	EL
0.1	0.0501	0.2846	0.2301	0.0522	0.5628	0.0494	0.8521	0.0594	0.9729	0.0660
0.2	0.0500	0.2306	0.2349	0.0554	0.5753	0.0582	0.8631	0.0604	0.9767	0.0874
0.3	0.0500	0.1864	0.2385	0.0640	0.5855	0.0764	0.8722	0.0886	0.9799	0.1164
0.4	0.0498	0.1608	0.2406	0.0608	0.5920	0.0794	0.8782	0.1008	0.9819	0.1486
0.5	0.0498	0.1510	0.2413	0.0762	0.5942	0.1048	0.8803	0.1380	0.9826	0.2000

Notes to table:

The number of replications is 5,000.

$C_0 = (\frac{\sigma_1^2}{n_1})(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})^{-1}$, $C_0 = \{.1, .2, .3, .4, .5\}$ corresponds to $\rho^2 = \{9, 4, 2.33, 1.5, 1\}$.

$\delta_0 = (\mu_1 - \mu_2)(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})^{-1} = \delta\sqrt{7}$. $\delta_0 = \{1, 2, 3, 4\}$, which is equivalent to $\delta = \{.378, .7559, 1.1339, 1.5119\}$.

Table 6: Size and Power of the ELR Test for $n_1 = n_2 = 7$

ρ^2	α	$\delta = 0$	$\delta = 0.3780$	$\delta = 0.7559$	$\delta = 1.1334$	$\delta = 1.5119$
1	10%	0.2014	0.1030	0.0992	0.1266	0.1612
	5%	0.1510	0.0522	0.0494	0.0594	0.0660
	2%	0.1072	0.0184	0.0194	0.0210	0.0218
	1%	0.0850	0.0082	0.0098	0.0082	0.0070
1.5	10%	0.2062	0.1134	0.1232	0.1274	0.1770
	5%	0.1608	0.0554	0.0582	0.0604	0.0874
	2%	0.1182	0.0204	0.0224	0.0208	0.0332
	1%	0.0966	0.0106	0.0112	0.0088	0.0102
2.33	10%	0.2338	0.1184	0.1398	0.1562	0.2058
	5%	0.1864	0.0640	0.0764	0.0886	0.1164
	2%	0.1400	0.0206	0.0282	0.0276	0.0374
	1%	0.1110	0.0094	0.0158	0.0106	0.0150
4	10%	0.2838	0.1216	0.1500	0.1756	0.2672
	5%	0.2306	0.0608	0.0794	0.1008	0.1486
	2%	0.1686	0.0278	0.0370	0.0434	0.0566
	1%	0.1366	0.0110	0.0168	0.0172	0.0244
9	10%	0.3384	0.1438	0.1974	0.2652	0.3754
	5%	0.2846	0.0762	0.1048	0.1380	0.2000
	2%	0.2140	0.0242	0.0386	0.0362	0.0508
	1%	0.1760	0.0126	0.0174	0.0180	0.0240

Notes to table:

The number of replications is 5,000.

$C_0 = \{.1, .2, .3, .4, .5\}$ corresponds to $\rho^2 = \{9, 4, 2.33, 1.5, 1\}$.

$\delta_0 = \delta\sqrt{7}$. $\delta_0 = \{1, 2, 3, 4\}$, which is equivalent to $\delta = \{.378, .7559, 1.1339, 1.5119\}$

References

- Aptech Systems, 2002. Gauss 5.0 for Windows NT, (Aptech Systems, Inc., Maple Valley WA).
- Aspin, A. A., 1948. "An Examination and Further Development of a formula Arising in the Problem of Comparing Two Mean Values", *Biometrika* 35, 88-96.
- Bera, A., Biliias, Y., 2002. "The MM, ME, ML, EL, EF and GMM Approaches to Estimation: A Synthesis", *Journal of Econometrics* 107, 51-86.
- Cochran, W. G., Cox, G. M., 1950. Experiment Designs (John Wiley and Sons, New York).
- Dong, L.B., Giles, D.E.A., 2004. "An empirical likelihood ratio test for normality", Working paper EWP0401, Department of Economics, University of Victoria.
- Fisher, R. A., 1935. "The Fiducial Argument in Statistical Inference", *Annals of Eugenics* 6, 391-198.
- Fisher, R. A., 1941. "The Asymptotic Approach to Behrens' Integral with further Tables for the d Test of Significance", *Annals of Eugenics* 11, 141-172.
- Giles, J. A., Giles, D. E. A., 1993. "Pre-Testing Estimation and Testing in Econometrics: Recent Developments", *Journal of Economic Surveys* 7, 145-197.
- Jing, B. Y., 1995. "Two-Sample Empirical Likelihood Method", *Statistics and Probability Letters* 24, 315-319.
- Lee, A. F. S., Gurland, J., 1975. "Size and Power of Tests for Equality of Means of Two Normal Populations", *Journal of the American Statistical Association* 70, 933-944.
- Mittelhammer, R., Judge, G., Miller, D., 2000. *Econometric Foundations* (Cambridge Uni-

versity Press, Cambridge).

Mittelhammer, R., G. Judge, and R. Schoenberg (2003), "Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods", *Working Paper*, Washington State University, University of California, Berkeley, and Aptech Systems, Inc.

Nelder, J. A., Mead, R., 1965. "A Simplex Method for Function Minimization", *Computer Journal* 7, 308-313.

Owen, A. B., 2001. *Empirical Likelihood* (Chapman & Hall/CRC, New York).

Qin, J., 1991. "Likelihood and Empirical Likelihood Ratio Confidence Intervals in Two Sample Semi-parametric Models", *Technical Report Series University of Waterloo Stat-91-6*.

Qin, J., 1993. "Empirical Likelihood in Biased Sample Problems", *Annals of Statistics* 21, 1182-1196.

Tukey, J. W., 1954. "Unsolved Problems of Experimental Statistics", *Journal of the American Statistical Association* 49, 706-731.

Weerahandi, S., 1987. "Testing Regression Equality with Unequal Variances", *Econometrica* 55, 1211-1215.

Welch, B. L., 1947. "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved", *Biometrika* 34, 28-35.

Zaman, A., 1996. *Statistical Foundations for Econometric Techniques* (Academic Press, New York).